

WHAT IS CLAIMED IS:

1. A method of generating a wrapper for extracting information from a repository of semistructured information, the method comprising the computer-implemented steps of:

presenting to a user a set of predefined operation types from which to construct the wrapper, the predefined set comprising operation types for extracting data from the repository based on structural relationships within the semistructured information;

presenting to the user an option to select operation types from the set and to specify properties of selected operation types, to allow the user to define operations to be performed by the wrapper;

monitoring and recording selections by the user of operation types and entry of operation properties, to generate the wrapper; and

generating and displaying on a display screen a graphical representation of the wrapper in which wrapper operations are depicted as objects on the display screen.

2. The method of Claim 1, further comprising presenting to the user an option to select operations depicted on the display screen to allow the user to define links between operations, wherein a link represents a call of a destination operation by a source operation.

3. The method of Claim 2, wherein links are graphically depicted on the display screen.

4. The method of Claim 3, wherein the repository comprises a Web site.

5. The method of Claim 4, wherein the set of predefined operation types provides the capability of matching a regular expression.

6. The method of Claim 5, wherein the generated wrapper is configured to output extracted data in a structured format.

7. The method of Claim 6, further comprising visually depicting execution of the wrapper through the graphical representation of the wrapper.

8. The method of Claim 7, further comprising displaying output of the wrapper.

9. The method of Claim 8, further comprising displaying data extracted from the semistructured information by the wrapper.

10. The method of Claim 9, further comprising performing a statistical analysis on the output of the wrapper.

5 11. A method of generating a wrapper for extracting data from a Web site, the method comprising:

invoking an application, the application providing a graphical design environment in which the wrapper can be graphically represented;

10 using the application to create objects in the graphical design environment, the objects representing operations for extracting data from a Web site based on structural relationships within the Web site; and

using the application to create links between the objects in the graphical design environment, wherein a link represents a call of a destination operation by a source operation, and wherein the wrapper comprises the operations and the calls.

15 12. The method of Claim 11, further comprising using the application to debug the wrapper.

13. The method of Claim 12, further comprising:

using the application to execute the wrapper; and

20 using the application to display the data extracted from the Web site in a structured format.

14. The method of Claim 13, further comprising using the application to examine the Web site to identify information of interest.

25 15. The method of Claim 11, further comprising defining a structured format in which the wrapper can output the information.

16. A computer program for facilitating the design and creation of a wrapper, the computer program comprising, on a computer readable medium:

wrapper editor code which provides a graphical design environment in which a wrapper can be graphically constructed;

30 wrapper rendering code which displays a graphical representation of the wrapper within the graphical design environment as the wrapper is constructed; and

operation class code which defines a set of operation classes for extracting data from semistructured information based on structural relationships within the semistructured information, wherein the wrapper comprises instances of the operation classes.

5 17. The computer program of Claim 16, wherein the semistructured information comprises a Web site.

18. The computer program of Claim 17, further comprising table display code which displays information extracted from the Web site.

10 19. The computer program of Claim 18, further comprising wrapper run code which enables the execution of the wrapper.

20. The computer program of Claim 19, further comprising wrapper debugging code.

21. The computer program of Claim 20, wherein the set of operation classes enable the structuring of extracted data.

15 22. The computer program of Claim 21, wherein the wrapper debugging code provides options for controlling the execution of the wrapper, the options comprising starting the wrapper, stopping the wrapper, and setting breakpoints.

23. The computer program of Claim 22, wherein the wrapper debugging code provides options for displaying transitory data produced by the wrapper during execution.

20 24. The computer program of Claim 22, further comprising Web viewer code which displays Web pages.

25 25. A method of executing a wrapper, the method comprising:
creating an instance of a wrapper within an internal memory of a
computer, the wrapper facilitating the automated extraction of information from
a Web site;

providing a graphical representation of the wrapper on a display screen;
providing a user interface through which a user can control the execution
of the wrapper;

30 monitoring the user interface for user input;
controlling the execution the wrapper in response to user input;

graphically depicting the execution of the wrapper through the graphical representation of the wrapper; and

displaying information extracted from the Web site.

26. The method of Claim 25, further comprising:

5 graphically depicting objects on the display screen, the objects representing operations to be performed by the wrapper; and graphically depicting links between objects on the display screen, wherein a link represents a call of a destination operation by a source operation.

10 27. The method of Claim 26, wherein the user interface provides to the user options for controlling the execution of the wrapper, the options comprising starting the wrapper, stopping the wrapper, and setting breakpoints at operations.

28. The method of Claim 27, further comprising performing a statistical analysis on output of the wrapper.

15 29. The method of Claim 27, wherein the wrapper is configured to output extracted information in a structured format.

Ad
17